

DIALOG(R) File 351: Derwent WPI  
(c) 2001 Derwent Info Ltd. All rights reserved.

008938876 \*\*Image available\*\*

WPI Acc No: 1992-066145/199209

XRPX Acc No: N92-049667

Speech interval establishment unit for speech recognition system -  
operates in two stages on filtered, multiplexed and digitised signals  
from speech and background noise microphones

Patent Assignee: RICOH KK (RICO )

Inventor: ARIYOSHI T

Number of Countries: 002 Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
DE 4126902	A	19920220	DE 4126902	A	19910814	199209 B
DE 4126902	C2	19960627	DE 4126902	A	19910814	199630
JP 3118023	B2	20001218	JP 91166391	A	19910611	200102

Priority Applications (No Type Date): JP 91124953 A 19910425; JP 90215196 A  
19900815; JP 90215197 A 19900815

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
DE 4126902	C2	10	G10L-005/00		
JP 3118023	B2	9	G10L-011/02		Previous Publ. patent JP 5006193

Abstract (Basic): DE 4126902 A

The signals from two microphones (1,2) at unequal distances from the speaker are processed by separate amplifiers (11,21), band-pass filters (12,22), rectifying circuits (13,23), low-pass filters (14,24), multiplexers (15,25) and A/D convertors (16,26).

An interval recognition unit (30) operates on the difference between the aggregate powers of the acoustic signal (X(i)) and the spectrum (N(i)) of a periodic ref. signal in a preset time frame. A noise coefft. calculator (40) and canceller (50) produce a speech spectrum (S(i)) for threshold comparison in a second stage of recognition (60).

USE/ADVANTAGE - In noisy environment, e.g. in , factory, home. Speech or voice intervals can be recognised adequately. even in extreme conditions of irregular noise. (12pp Dwg.No. 1/4)

Title Terms: SPEECH; INTERVAL; ESTABLISH; UNIT; SPEECH; RECOGNISE; SYSTEM; OPERATE; TWO; STAGE; FILTER; MULTIPLEX; DIGITAL; SIGNAL; SPEECH; BACKGROUND; NOISE; MICROPHONE

Derwent Class: P86; W04

International Patent Class (Main): G10L-005/00; G10L-011/02

International Patent Class (Additional): G10L-015/02; G10L-015/04;

G10L-015/20; G10L-021/02; G10L-101-023

File Segment: EPI; EngPI

Manual Codes (EPI/S-X): W04-V04A; W04-V04A1; W04-V05E

19 BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENTAMT

12 Patentschrift  
10 DE 41 26 902 C 2

61 Int. Cl.<sup>8</sup>:  
G 10 L 5/00

21 Aktenzeichen: P 41 26 902.0-53  
22 Anmeldetag: 14. 8. 91  
43 Offenlegungstag: 20. 2. 92  
45 Veröffentlichungstag  
der Patenterteilung: 27. 6. 96

DE 41 26 902 C 2

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

30 Unionspriorität: 32 33 31

15.08.90 JP 2-215196 15.08.90 JP 2-215197  
25.04.91 JP 3-124953

73 Patentinhaber:

Ricoh Co., Ltd., Tokio/Tokyo, JP

74 Vertreter:

Schwabe, H., Dipl.-Ing.; Sandmair, K., Dipl.-Chem.  
Dr.jur. Dr.rer.nat.; Marx, L., Dipl.-Phys. Dr.rer.nat.,  
Pat.-Anwälte, 81677 München

72 Erfinder:

Ariyoshi, Takashi, Yokohama, Kanagawa, JP

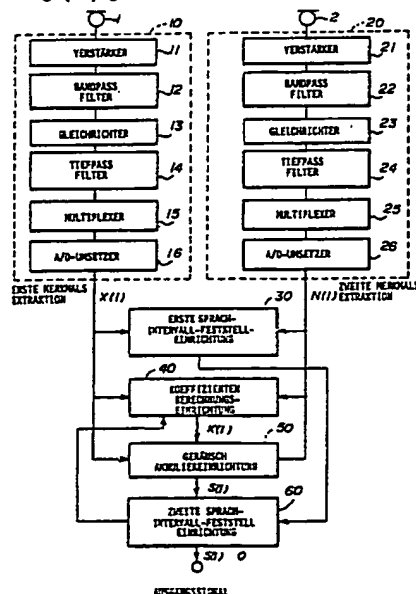
66 Für die Beurteilung der Patentfähigkeit  
in Betracht gezogene Druckschriften:

DE 38 42 591 C1  
DE 15 12 758 C1  
DE 38 37 066 A1  
US 49 12 767  
EP 03 32 890 A2  
JP 63-2 62 895  
JP 58-1 86 599  
JP 58-1 30 395  
JP 56-1 35 898  
JP 01-2 39 596

54 Sprachintervall - Feststelleinheit

57 Sprachintervall-Feststelleinheit, um ein Sprachintervall, das einer Sprache oder Stimme eines Sprechers entspricht, in einem Tonsignal festzustellen, mit einer ersten Eingabeeinrichtung (1), welche nahe bei einem Sprecher vorgesehen ist, um Töne einschließlich einer Stimme oder Sprache des Sprechers und Geräusche in der Umgebung einzugeben und um ein Tonsignal, das den eingegebenen Tönen entspricht, abzugeben; mit einer zweiten Eingabeeinrichtung (2), die weit von dem Sprecher weg vorgesehen ist, um Töne einschließlich Geräusche in der Umgebung einzugeben und um ein Bezugssignal abzugeben, das den eingegebenen Tönen entspricht; mit einem Bandpaßfilter zur Aufteilung des Tonsignals in eine Anzahl vorherbestimmter Frequenzbänder; mit einer ersten Merkmalsextraktionseinrichtung (10), welche mit der ersten Eingabeeinrichtung (1) verbunden ist, um ein erstes Merkmal in jedem der Frequenzbänder aus dem Tonsignal zu extrahieren, das von der ersten Eingabeeinrichtung (1) abgegeben worden ist; mit einer zweiten Merkmalsextraktionseinrichtung (20), welche mit der zweiten Eingabeeinrichtung (2) verbunden ist, um ein zweites Merkmal in jedem der Frequenzbänder aus dem Bezugssignal zu extrahieren, das von der zweiten Eingabeeinrichtung (2) abgegeben worden ist; mit einer Koeffizientenberechnungseinrichtung (40), welche mit der ersten und der zweiten Merkmalsextraktionseinrichtung (10, 20) verbunden ist, um einen Geräuschkoeffizienten in jedem der Frequenzbänder basierend auf dem ersten und dem zweiten Merkmal zu berechnen, wenn keine Sprache des Sprechers vorliegt; mit einer Geräuschnulliereinrichtung (50), welche mit der ersten und zweiten Merkmalsextraktionseinrichtung (10, 20) und der Koeffizientenberechnungseinrichtung (40) verbunden ist, um ein Sprachintervall in jedem der Frequenzbänder basierend auf dem ersten und zweiten Merkmal und dem Geräuschkoeffizienten durch Eliminierung von Geräuschkomponenten aus dem ersten Merkmal abzuleiten;

mit einer ersten Sprachintervall-Feststelleinrichtung (30) zum Feststellen eines generellen Sprachintervalls, welches der Sprache oder Stimme des Sprechers in dem akustischen Signal entspricht, und mit einer zweiten Sprachintervall-Feststelleinrichtung (60), welche mit der ersten Sprachintervall-Feststelleinrichtung (30) und der Geräuschnulliereinrichtung (50) verbunden ist, um ein Sprachintervall, in jedem der Frequenzbänder in einem Feststellintervall auf Grund des von der Geräuschnulliereinrichtung (50) gelieferten Merkmal festzustellen, ...



DE 41 26 902 C 2

Die Erfindung betrifft eine Sprachintervall-Feststellereinheit nach dem Anspruch 1.

Aus der DE 38 37 066 A1 ist bereits eine Sprachintervall-Feststellereinheit bekannt, um ein Sprachintervall, das einer Sprache bzw. Stimme eines Sprechers entspricht, in einem aufgenommenen Tonsignal festzustellen. Es ist eine erste Eingabeeinrichtung nahe bei einem Sprecher aufgestellt, um Töne einschließlich einer Stimme bzw. Sprache des Sprechers gemischt mit Geräuschen in der Umgebung aufzunehmen, und um ein Tonsignal, das den eingegebenen Tönen entspricht, zu erzeugen. Eine zweite Eingabeeinrichtung ist entfernt vom Sprecher aufgestellt, um Töne einschließlich von Geräuschsignalen in der Umgebung aufzunehmen und um ein Bezugssignal zu erzeugen, das den eingegebenen Tönen entspricht. Diese bekannte Sprachintervall-Feststellereinheit umfaßt ferner eine erste Merkmalsextraktionseinrichtung, um ein erstes Merkmal aus dem Tonsignal der ersten Eingabeeinrichtung zu extrahieren, und umfaßt eine zweite Merkmalsextraktionseinrichtung, um ein zweites Merkmal aus dem genannten Bezugssignal zu extrahieren, das von der zweiten Eingabeeinrichtung geliefert wird. Mit Hilfe einer Koeffizientenberechnungseinrichtung, welche die Ausgangssignale der zwei Merkmalsextraktionseinrichtungen empfängt, werden Geräuschkoeffizienten berechnet und es wird dann mit Hilfe einer Geräuschannulliereinrichtung eine Geräuschkomponente beseitigt, und zwar auf der Grundlage der berechneten Koeffizienten und dem ersten und zweiten extrahierten Merkmal.

Diese bekannte Sprachintervall-Feststellereinheit ist aber nicht dafür ausgebildet, um mit hoher Genauigkeit und Sicherheit Sprachintervalle innerhalb einer Anzahl von Frequenzbändern zu ermitteln.

In einem üblichen Spracherkennungssystem ist eine Geräuschbeseitigung ein wichtiges Thema. Hierbei ist es schwierig, Sprachintervalle, die Sprachabschnitten einer sprechenden Person entsprechen, eindeutig in einem akustischen Signal festzustellen, welchem Geräusche überlagert sind. Ein Verfahren, um Sprachintervalle in einem akustischen Signal festzustellen, welchem Geräusche überlagert sind, ist beispielsweise in der japanischen Offenlegungsschrift 56-135898, die der japanischen Patentschrift Nr. 63-29754 entspricht, beschrieben. In diesem Verfahren werden zwei Schwellenwertpegel verwendet, welche ein erster Schwellenwertpegel und ein zweiter Schwellenwertpegel sind, welcher niedriger als der erste Schwellenwertpegel ist, um die Sprachintervalle festzustellen. Das heißt, in einem Fall, bei welchem der Pegel des tonfrequenten Signals für eine vorherbestimmte Zeit oder länger ständig höher als der erste Schwellenwert ist, wird ein Intervall, in welchem der Pegel des tonfrequenten Signals höher als der zweite Schwellenwertpegel oder gleich diesem ist, als das Sprachintervall festgestellt. Jedoch ist es bei diesem herkömmlichen Verfahren in einem Fall, bei welchem der Pegel eines unregelmäßigen Geräusches, welches unregelmäßig erzeugt ist, hinsichtlich der Zeitdauer größer als der erste Schwellenwert ist, schwierig, eindeutig die Sprachintervalle festzustellen.

Ein Verfahren, um Sprachintervalle unter Bedingungen festzustellen, bei welchem unregelmäßige Geräusche dem akustischen Signal überlagert sind, ist beispielsweise in der offengelegten japanischen Offenlegungsschrift Nr. 58-130 395 beschrieben. Bei diesem Verfahren werden ein erstes Mikrophon für Stimmen

und ein zweites Mikrophon für Geräusche verwendet, um die Stimm- bzw. Sprachintervalle festzustellen. Das heißt, das Sprachintervall wird auf der Basis einer Differenz zwischen einer ersten Leistung eines Ausgangssignals von dem ersten Mikrophon und einer zweiten Leistung eines Ausgangssignals von dem zweiten Mikrophon festgestellt. Jedoch können bei diesem herkömmlichen Verfahren Geräuschkomponenten, welche in dem Sprachintervall enthalten sind, nicht beseitigt werden.

Außerdem ist ein spektrales Subtraktionsverfahren als ein Verfahren bekannt, um Geräuschkomponenten aus dem tonfrequenten Signal zu beseitigen (z. B. DE 36 42 591 C1). Jedoch ist es bei dem spektralen Subtraktionsverfahren schwierig, unregelmäßige Geräusche aus dem Tonsignal zu beseitigen. Bei einem adaptiven Geräuschannullierverfahren, das in den japanischen Offenlegungsschriften Nr. 58-196 599, 63-262 695 und 1-239 596 (welche dem US-Patent Nr. 4 912 767 entspricht) beschrieben ist, können die unregelmäßigen Geräusche wirksam aus dem Tonsignal entfernt werden in dem adaptiven Geräuschannullierverfahren werden die Geräuschkomponenten aus dem Tonsignal mit Hilfe einer ersten Toneingabeeinheit (einem Mikrophon) und einer zweiten Toneingabeeinheit (Mikrophon) wie folgt beseitigt.

Die erste Toneingabeeinheit ist nahe bei dem Sprecher vorgesehen, während die zweite Eingabeeinheit weit weg von dem Sprecher vorgesehen ist. Sprachabschnitte des Sprechers werden hauptsächlich in die erste Eingabeeinheit eingegeben. Verschiedene Arten von Geräuschen in der Umgebung werden sowohl in die erste als auch in die zweite Eingabeeinheit eingegeben. Ein Spektrum  $X(i)$  in jeweils einer Anzahl Frequenzbänder wird aus einem Tonsignal erhalten. Ein Spektrum  $N(i)$  in jedem der Frequenzbänder wird aus einem Signal erhalten, das von der zweiten Eingabeeinheit zugeführt worden ist. In diesem Fall ist mit  $i$  das  $i$ -te Frequenzband bezeichnet. Ein Geräuschverhältnis  $k(i)$  in jedem Frequenzband ist vorher entsprechend der folgenden Formel unter der Voraussetzung berechnet worden, daß es nicht die Stimme des Sprechers ist:

$$k(i) = X(i)/N(i).$$

Das Geräuschverhältnis  $k(i)$  stellt ein Verhältnis eines Geräusches, das von der ersten Eingabeeinheit festgestellt worden ist, zu einem Geräusch dar, das von der zweiten Eingabeeinheit in dem jeweiligen Frequenzband festgestellt worden ist.

Es wird dann ein Schätzwert  $S(i)$  eines Sprachspektrums, welches der Sprache bzw. Stimme des Sprechers entspricht, entsprechend der folgenden Formel berechnet:

$$S(i) = X(i) - k(i) \cdot N(i)$$

In diesem adaptiven Geräuschannullierverfahren ändert sich in einem Fall, bei welchem eine Geräuschquelle bezüglich eines  $i$ -ten Frequenzbandes vorhanden ist, obwohl sich der Pegel des Geräusches ändert, das Geräuschverhältnis  $k(i)$  nicht. Folglich kann eine Geräuschkomponenten, welche in dem Spektrum  $X(i)$  des Tonsignals in dem Sprachintervall enthalten ist, durch  $k(i) N(i)$  geschätzt werden. Das heißt, mit dem adaptiven Geräuschannullierverfahren kann das unregelmäßige Geräusch aus der akustischen Komponente beseitigt werden. Wenn jedoch in dem adaptiven Geräuschannullier-

verfahren das Geräuschverhältnis  $k(i)$  unter der Voraussetzung berechnet wird, daß der Pegel des Geräusches verhältnismäßig niedrig ist, wird ein Fehler des Schätzwertes  $S(i)$  des Sprachspektrums groß. Außerdem kann, wenn der Pegel des in dem Sprachintervall enthaltenen Geräusches verhältnismäßig hoch ist, das Sprach- bzw. Stimmenspektrum nicht angemessen geschätzt werden. Folglich können mit dem adaptiven Geräuschannullierverfahren die Geräuschkomponenten nur in einem Fall angemessen beseitigt werden, bei welchem Geräuschkomponenten immer in allen Frequenzbändern vorhanden sind.

Außerdem ist in DE 15 12 758 C1 ein Verfahren beschrieben, um den Schätzwert  $S(i)$  des Sprachspektrums entsprechend der folgenden Formel zu berechnen:

$$S(i) = X(i) - N(i).$$

Die vorstehende Formel enthält jedoch nicht den Koeffizienten  $k(i)$ . Folglich ergibt dieses Verfahren nicht das adaptive Geräuschannullierverfahren und es hat den Nachteil, daß der Schätzwert  $S(i)$  in dem Fall nicht korrekt ist, daß Geräuschquellen nahe bei dem Mikrophon vorhanden sind.

Die der Erfindung zugrundeliegende Aufgabe besteht darin, eine verbesserte Sprachintervall-Feststelleinheit zu schaffen, die insbesondere auch bei stark geräuschbelasteten Bedingungen, bei denen auch sehr unregelmäßige Geräusche auftreten können, eine hochgenaue Sprachintervall-Feststellung ermöglicht.

Diese Aufgabe wird erfindungsgemäß durch die im Anspruch 1 aufgeführten Merkmale gelöst.

Besonders vorteilhafte Ausgestaltungen und Weiterbildungen der Erfindung ergeben sich aus den Unteransprüchen.

Im folgenden wird die Erfindung anhand von bevorzugten Ausführungsbeispielen unter Hinweis auf die Zeichnungen näher erläutert. Es zeigt

Fig. 1 ein Blockdiagramm einer Sprachintervall-Feststelleinheit gemäß einer Ausführungsform mit Merkmalen nach der Erfindung;

Fig. 2 ein Zeitdiagramm, in welchem Beispiele von Geräuschintervallen wiedergegeben sind, welche mittels der in Fig. 1 dargestellten Sprachintervall-Feststelleinheit festgestellt werden;

Fig. 3 ein Blockdiagramm einer Geräuschannuliereinheit, welche in einem Spracherkennungssystem mit Merkmalen nach der Erfindung verwendet ist; und

Fig. 4 ein Blockdiagramm eines Spracherkennungssystems gemäß einer Ausführungsform mit Merkmalen nach der Erfindung.

Nachstehend wird eine bevorzugte Ausführungsform mit Merkmalen nach der Erfindung anhand von Fig. 1 und 2 beschrieben. In Fig. 1 ist eine Sprachintervall-Feststelleinheit dargestellt, welche eine erste Eingabe-einrichtung (Mikrophon) 1, eine zweite Eingabeeinrichtung (Mikrophon) 2, eine erste und eine zweite Merkmalsextraktionseinrichtung (Merkmalsextraktionsblock) 10 und 20, eine erste Sprachintervall-Feststelleinrichtung (Sprachintervall-Feststellblock) 30 eine Koeffizientenberechnungseinrichtung (Koeffizienten-Berechnungsblock) 40, eine Geräuschannuliereinrichtung (Geräuschannullierblock) 50 und eine zweite Sprachintervall-Feststelleinrichtung (Sprachintervall-Feststellblock) 60 aufweist. Das erste Mikrophon 1 ist nahe bei einem Sprecher vorgesehen, während das zweite Mikrophon 2 weit von dem Sprecher weg vorgesehen ist. Folglich wird eine Sprecherstimme oder -sprache

hauptsächlich in dem ersten Mikrophon 1 aufgenommen. Geräusche in der Umgebung können sowohl von dem ersten als auch von dem zweiten Mikrophon 1 bzw. 2 aufgenommen werden. Wenn es die Stimme des Sprechers ist, gibt das erste Mikrophon 1 ein T-signal ab, welches Komponenten, welche der Stimme des Sprechers entsprechen, und Komponenten enthält, welche den Geräuschen in der Umgebung entsprechen. Das zweite Mikrophon 2 gibt ein Bezugssignal ab, das Geräuschen in der Umgebung entspricht. Das Bezugssignal, das von dem zweiten Mikrophon 2 abgegeben wird, enthält kaum eine Komponente der Stimme bzw. Sprache des Sprechers.

Der erste Merkmalsextraktionsblock 10 extrahiert Merkmalsdaten aus dem von dem ersten Mikrophon 1 gelieferten Tonsignal. Der erste Merkmalsextraktionsblock 10 hat einen Verstärker 11, ein Bandpaßfilter 12, einen Gleichrichter 13, ein Tiefpaßfilter 14, einen Multiplexer 15 und einen Analog-Digital-Umsetzer 16. Der Verstärker 11 verstärkt das von dem ersten Mikrophon 1 gelieferte Tonsignal mit einem vorherbestimmten Verstärkungsfaktor. Das Bandpaßfilter 12 hat fünfzehn Kanäle, die jeweils einem von fünfzehn Frequenzbändern entsprechen. Das Bandpaßfilter 12 teilt das Tonsignal in ein Spektrum auf, welches aus fünfzehn Frequenzbändern besteht. Jeder Wert der fünfzehn Frequenzbänder wird von dem Gleichrichter 13 und dem Tiefpaßfilter 14 verarbeitet, so daß ein Spektrum  $X(i)$  des Tonsignals in den fünfzehn Frequenzbändern (Kanälen) erhalten wird. Der Multiplexer 15 wählt eines der fünfzehn Frequenzbänder aus, welches Spektrum von dem Bandpaßfilter 12 über den Gleichrichter 13 und das Tiefpaßfilter 14 geliefert wird. Der Analog-Digital-Umsetzer 16 setzt das Spektrum  $X(i)$  in jedem Frequenzband in digitale Daten um. Folglich gibt der erste Merkmalsextraktionsblock 10 periodisch das Spektrum  $X(i)$  ( $i = 1, 2, \dots, 15$ ) des akustischen Signal bei einer vorherbestimmten Rahmenzeit ab.

Der zweite Merkmalsextraktionsblock 20 hat einen Verstärker 21, ein Bandpaßfilter 22, einen Gleichrichter 23, ein Tiefpaßfilter 24, einen Multiplexer 25 und einen Analog-Digital-Umsetzer 26 genauso wie der erste Merkmalsextraktionsblock. Folglich gibt der zweite Merkmalsextraktionsblock 20 periodisch ein Spektrum  $N(i)$  des Referenzsignals bei der vorherbestimmten Rahmenzeit ab. Ein Verstärkungsfaktor des Verstärkers 11 in dem ersten Merkmalsextraktionsblock 10 und ein Verstärkungsfaktor des Verstärkers 21 in dem zweiten Merkmalsextraktionsblock 20 werden so eingestellt, daß der Pegel des Tonsignals annähernd gleich dem Pegel des Referenzsignals wird. Kenndaten anderer Teile in dem ersten Merkmalsextraktionsblock 10 sind gleich Kenndaten entsprechender Teile in dem zweiten Merkmalsextraktionsblock 20. Die beiden Merkmalsextraktionsblöcke 10 und 20 können auch Merkmalsdaten außer den vorerwähnten Spektren  $X(i)$  und  $N(i)$  extrahieren.

Der erste Sprachintervall-Feststellblock 30 stellt ein Sprachintervall basierend auf einer Differenz zwischen einer Gesamtleistung  $\Sigma X(i)$  des akustischen Signals und einer Gesamtleistung  $\Sigma N(i)$  des Bezugssignals fest. Der erste Sprachintervall-Feststellblock 30 stellt das Sprachintervall fest, wenn die Differenz zwischen der Gesamtleistung  $\Sigma X(i)$  des akustischen Signals und der Gesamtleistung  $\Sigma N(i)$  größer als ein Schwellenwert  $T_{\text{PWT}}$  ist. Auf das mittels des ersten Spannintervall-Feststellblocks 30 festgestellte Sprachintervall wird als ein generelles Sprachintervall Bezug genommen. Der

Schwellenwert  $T_{pwr}$  wird auf der Basis eines Mittelwerts  $Av X_{pwr}$  der Gesamtleistung des Tonsignals in einer Anzahl Rahmen  $v$  r dem gegenwärtigen Rahmen und in einem Intervall außer dem generellen Sprachintervall und auf der Basis eines Mittelwerts  $Av N_{pwr}$  der Gesamtleistung des Bezugssignals berechnet, das unter derselben Bedingung wie der Mittelwert  $Av X_{pwr}$  erhalten worden ist. Das heißt, der Schwellenwert  $T_{pwr}$  wird berechnet und entsprechend der folgenden Formel (1) aktualisiert:

$$T_{pwr} = a_{pwr} \cdot (Av X_{pwr} - Av N_{pwr}) + b_{pwr} \quad (1)$$

wobei  $a_{pwr}$  und  $b_{pwr}$  konstante Zahlen und größer als null sind ( $a_{pwr}, b_{pwr} > 0$ ).

Der Geräuschannullierblock 50 erzeugt ein Merkmal  $S(i)$  der Sprache, aus welchem Geräuschkomponenten beseitigt werden, wie später noch beschrieben wird. Eine Sprachleistung  $\Sigma S(i)$  wird auf der Basis des Merkmals  $S(i)$  der Sprache erhalten, und dann kann das generelle Sprachintervall auf der Basis eines Vergleichs der Sprachleistung  $\Sigma S(i)$  mit dem Schwellenwert  $T_{pwr}$  erhalten werden.

Der Koeffizientenberechnungsblock 40 berechnet einen Geräuschkoeffizienten  $k(i)$  in jedem Kanal entsprechend der folgenden Formel (2), welche das Spektrum  $X(i)$  des Tonsignals und das Spektrum  $N(i)$  des Bezugssignals enthält, welche unter der Bedingung erhalten werden, daß keine Sprache des Sprechers vorhanden ist:

$$k(i) = [X(i) + C_1] / [N(i) + C_2] \quad (2)$$

wobei  $C_1$  und  $C_2$  konstante Zahlen und größer als null sind ( $C_1, C_2 > 0$ ).

Der Geräuschkoeffizient  $k(i)$  kann auch auf der Basis eines Mittelwerts der Spektren  $x(i)$  und  $N(i)$  in einer Anzahl Rahmen vor dem gegenwärtigen Rahmen berechnet werden. In diesem Fall ist die Anzahl Rahmen, um den Mittelwert zu erhalten, vorzugsweise kleiner als die Anzahl Rahmen, die erforderlich sind, um den vorstehenden Schwellenwert  $T_{pwr}$  und einen Schwellenwert  $T_1$  zu berechnen, welcher später noch beschrieben wird.

Gemäß der vorstehenden Formel (2) gilt: je größer die Spektren  $X(i)$  und  $N(i)$  sind, umso näher bzw. dichter kommt Verhältnis  $X(i)/N(i)$ , welches das herkömmliche Geräuschverhältnis ist, der Geräuschkoeffizient  $k(i)$ . Je kleiner die Spektren  $X(i)$  und  $N(i)$  sind, umso näher dem Verhältnis  $C_1/C_2$  kommt andererseits der Geräuschkoeffizient  $k(i)$ . Folglich kann in einem Fall, bei welchem die Spektren  $x(i)$  und  $N(i)$  in jedem Frequenzband klein sind, wenn  $C_1/C_2$  auf einen angemessenen Wert in dem Spracherkennungssystem gesetzt ist, ein Fehler des Geräuschkoeffizienten  $k(i)$  klein werden. Falls die jeweiligen Werte von  $X(i)$  und  $N(i)$  durch 8 Bits (0 bis 255) dargestellt werden, sind jeweils  $C_1$  und  $C_2$  vorzugsweise Werte in einem Bereich zwischen 8 und 32. Falls der Pegel von  $X(i)$  und der Pegel von  $N(i)$ , welche unter einer Bedingung gemessen werden, bei welcher ein Geräusch, das von einer Quelle weit weg von den Mikrofonen 1 und 2 erzeugt worden ist, oder ein Geräusch vorhanden ist, das durch eine bestimmte Quelle erzeugt worden ist, annähernd einander gleich sind, kann der Wert von  $C_1/C_2$  auf "1" gesetzt werden. In anderen Fällen kann der Wert von  $C_1/C_2$  auf einen Wert von  $x(i)/N(i)$  gesetzt werden, wobei  $X(i)$  bzw.  $N(i)$  unter einer Bedingung gemessen werden, bei welcher keine Sprache des Sprechers vorliegt. Der Wert von  $C_1/C_2$  kann in

Abhängigkeit von dem jeweiligen Kanal auf einen entsprechenden Wert gesetzt werden. In einem System, in welchem eine Sprache des Sprechers an dem zweiten Mikrophon 2 eingegeben werden kann, ist, wenn der Wert von  $C_1/C_2$  gleich "1" ist ( $C_1 = C_2$ ), ein Teil der Sprachkomponenten als eine Geräuschkomponente aus der Sprache beseitigt. Folglich ist in diesem Fall der Wert von  $C_1/C_2$  vorzugsweise kleiner als "1" ( $C_1 < C_2$ ).

Der Wert, welcher durch Glätten des Geräuschkoeffizienten  $k(i)$  bezüglich einer Zeitskala erhalten worden ist, kann verwendet werden, um das Geräusch aus der Sprache zu eliminieren. Die Beziehung zwischen dem Geräuschkoeffizienten  $k(i)$  und den Spektren  $X(i)$  und  $N(i)$  ist nicht auf diejenige beschränkt, welche durch die vorstehende Formel (1) wiedergegeben ist. Die Beziehung kann auch durch eine hyperbolische Funktion, eine Exponentialfunktion u.ä. dargestellt werden.

Der Geräuschannullierblock 50 berechnet das Merkmal  $S(i)$  der Sprache, aus welchem das Geräusch beseitigt ist, in jedem Kanal (jedem Frequenzband) entsprechend der folgenden Formel (3):

$$S(i) = X(i) - k(i) \cdot N(i) \quad (3)$$

In einem Intervall außer dem Sprachintervall kann  $S(i)$  auf "0" gesetzt werden.

Der zweite Sprachintervall-Feststellblock 60 addiert vorherbestimmte Intervalle zu einem vorderen Ende des von dem ersten Sprachintervall-Feststellblock 30 festgestellten, generellen Sprachintervalls und zu einem hinteren Ende des generellen Sprachintervalls. Folglich erzeugt der zweite Sprachintervall-Feststellblock 60 ein Feststellintervall, welches aus dem generellen Sprachintervall und dem dazu addierten Intervall gebildet ist. Dann bestimmt der zweite Sprachintervall-Feststellblock 60, ob das Spektrum  $S(i)$  der Sprache, welche durch den Geräuschannullierblock 50 erhalten wird, größer als ein Schwellenwert  $T(i)$  in dem dabei erhaltenen Feststellintervall ist oder nicht. Wenn das Spektrum  $S(i)$  in dem  $i$ -ten Frequenzband der Sprache größer als der Schwellenwert  $T(i)$  ist, stellt der zweite Sprachfeststellblock 60 ein Sprachintervall in dem  $i$ -ten Kanal fest, welcher dem  $i$ -ten Frequenzband entspricht. Das heißt, der zweite Sprachintervall-Feststellblock 60 gibt das Spektrum  $S(i)$ , das von dem Geräuschannullierblock 50 geliefert worden ist, in dem Sprachintervall in jedem Frequenzband und "0" in einem Intervall außer dem Sprachintervall ab. Das Spektrum  $S(i)$  (einschließlich "0"), das von dem zweiten Sprachintervall-Feststellblock 60 abgegeben worden ist, wird in einem Prozeß zum Erkennen von Sprache eines Sprechers verwendet.

In Fig. 2 sind Sprachintervalle dargestellt, welche durch den ersten und zweiten Sprachintervall-Feststellblock 30 bzw. 60 festgestellt werden, wenn eine Sprache, wie beispielsweise "SECHS" an dem ersten Mikrophon 1 eingegeben wird. In Fig. 2 stellt der erste Sprachintervall-Feststellblock 30 das generelle Sprachintervall fest, was durch (A) in Fig. 2 dargestellt ist. Der zweite Sprachintervall-Feststellblock 60 addiert vorherbestimmte Intervalle zu den generellen Sprachintervallen, um so das Feststellintervall zu erzeugen, das durch (B) in Fig. 2 dargestellt ist. Dann stellt der zweite Sprachintervall-Feststellblock 60 das Sprachintervall in jedem der fünfzehn Frequenzbänder von einem niedrigen (dem ersten) Frequenzband bis zu einem hohen (dem fünfzehnten) Frequenzband in dem durch (B) dargestellten Feststellsprachintervall fest, was durch (C) in Fig. 2 dargestellt ist.

Der vorerwähnte Schwellenwert  $T_i$  welcher verwendet wird, um das Sprachintervall in jedem Frequenzband festzustellen, wird entsprechend der folgenden Formel (4) berechnet und aktualisiert:

$$T_i = a \cdot (A_v X(i) - A_v N(i)) + b \quad (4)$$

(a, b : konstante Zahlen, a, b > 0).

In der vorstehenden Formel (4) stellt  $A_v X(i)$  einen Mittelwert der Spektren  $x(i)$  des Tonsignals in einer Anzahl Rahmen vor dem gegenwärtigen Rahmen und in Intervallen außer dem Sprachintervall in dem i-ten Frequenzband dar;  $A_v N(i)$  stellt einen Mittelwert der Spektren  $N(i)$  des Bezugssignals in der Anzahl Rahmen vor dem gegenwärtigen Rahmen in dem Intervall außer dem Sprachintervall in dem i-ten Frequenzband dar.

Falls das in Fig. 1 dargestellte System in Umgebungen vorgesehen wird, in welchen Geräusche nicht zu groß sind, kann der Prozeß in dem ersten Sprachintervall-Feststellblock 30 weggelassen werden. Das heißt, der zweite Sprachintervall-Feststellblock 60 kann unmittelbar das Sprachintervall in jedem der Frequenzbänder in dem Spektrum  $X(i)$  des akustischen Signals feststellen. Der Schwellenwert  $T(i)$ , welcher dazu verwendet wird, das Sprachintervall in jedem der Frequenzbänder festzustellen, bzw. die Geräuschkoeffizienten  $k(i)$  können in einer Gruppe berechnet werden, welche eine Anzahl Kanäle (Frequenzbänder) enthält.

Der erste Sprachintervall-Feststellblock 30, der Koeffizienten-Berechnungsblock 40, der Geräuschannullierblock 50 und der zweite Sprachintervall-Feststellblock (60) sind beispielsweise in einem Computersystem gebildet.

In Fig. 3 ist eine weitere Ausführungsform mit Merkmalen nach der Erfindung dargestellt, wobei diejenigen Teile, welche dieselben wie in Fig. 1 sind, mit denselben Bezugszeichen versehen sind. In Fig. 3 ist das zweite, in Fig. 1 dargestellte Mikrophon 2 durch einen Lautsprecher 2s ersetzt. Das heißt, die Töne, welche von dem Lautsprecher 2s abgegeben worden sind, werden als Geräusche dem ersten Mikrophon 1 zugeführt. Ein Tonsignal, das dem Lautsprecher 2s zugeführt wird, wird ebenfalls als ein Geräuschsignal dem zweiten Merkmalsextraktionsblock 20 zugeführt. Der Sprachintervall-Feststellblock 60 stellt das Sprachintervall in jedem der Frequenzbänder basierend auf dem Schwellenwert  $T_i$  fest, welcher entsprechend der Formel (4) berechnet worden ist. Der Koeffizienten-Berechnungsblock 40 berechnet die Geräuschkoeffizienten  $k(i)$  entsprechend der vorstehenden Formel (2). Der Geräuschannullierblock 50 gibt das Merkmal  $S(i)$  (Spektrum) der Sprache ab, was entsprechend der Formel (3) berechnet worden ist. Die Verstärkungsfaktoren der Verstärker in den beiden Merkmalsextraktionsblöcken 10 und 20 werden so eingestellt, daß das akustische Signal und das Bezugssignal annähernd einander gleich sind, wenn das Tonsignal dem Lautsprecher 2s und dem zweiten Merkmalsextraktionsblock 20 zugeführt wird.

In dem in Fig. 3 dargestellten System kann das dem Tonsignal entsprechende Geräusch aus dem Tonsignal beseitigt werden, das von dem ersten Merkmalsextraktionsblock 10 erhalten worden ist.

In Fig. 4 ist ein Spracherkennungssystem gemäß einer Ausführungsform mit Merkmalen nach der Erfindung dargestellt. In Fig. 4 sind diejenigen Teile, welche dieselben Teile wie in Fig. 1 sind, mit denselben Bezugszeichen bezeichnet. In Fig. 4 hat das Spracherkennungssystem das erste Mikrophon 1, das zweite Mikrophon 2,

den Sprachintervall-Feststellblock 60, den Koeffizienten-Berechnungsblock 40 und den Geräuschannullierblock 50 in der gleichen Weise wie die in Fig. 1 und 3 dargestellten Systeme. Das Spracherkennungssystem hat auch einen Eingangsmustergenerator 70, einen Referenzmusterspeicher (Bezugsmusterspeicher) 80 und einen Erkennungsblock (Erkennungseinrichtung) 90. Der Eingangsmuster-Generator 70 erzeugt ein Eingangsmuster auf der Basis des Spektrums  $S(i)$  der Sprache, welche von dem Geräuschannullierblock 50 geliefert wird. Der Referenzmusterspeicher 80 speichert vorherbestimmte Referenzmuster von Sprachen. Der Erkennungsblock 90 vergleicht das Eingangsmuster mit den Referenzmustern in dem Referenzmusterspeicher 80 und führt einen Erkennungsprozeß durch. Ausführung und Arbeitsweise in dem Eingangsmuster-Generator 70, dem Bezugsmusterspeicher 80 und dem Erkennungsblock 90 sind beispielsweise diejenigen, welche in einem bekannten BTSP-Spracherkennungsverfahren verwendet sind.

In dem vorstehend beschriebenen Spracherkennungssystem wird ein Referenzmuster, welches dem Eingangsmuster angepaßt ist von dem Erkennungsblock 90 als das Erkennungsergebnis abgegeben.

#### Patentansprüche

1. Sprachintervall-Feststelleinheit, um ein Sprachintervall, das einer Sprache oder Stimme eines Sprechers entspricht, in einem Tonsignal festzustellen, mit
  - einer ersten Eingabeeinrichtung (1), welche nahe bei einem Sprecher vorgesehen ist, um Töne einschließlich einer Stimme oder Sprache des Sprechers und Geräusche in der Umgebung einzugeben und um ein Tonsignal, das den eingegebenen Tönen entspricht, abzugeben;
  - mit einer zweiten Eingabeeinrichtung (2), die weit von dem Sprecher weg vorgesehen ist, um Töne einschließlich Geräusche in der Umgebung einzugeben und um ein Bezugssignal abzugeben, das den eingegebenen Tönen entspricht;
  - mit einem Bandpaßfilter zur Aufteilung des Tonsignals in eine Anzahl vorherbestimmter Frequenzbänder;
  - mit einer ersten Merkmalsextraktionseinrichtung (10), welche mit der ersten Eingabeeinheit (1) verbunden ist, um ein erstes Merkmal in jedem der Frequenzbänder aus dem Tonsignal zu extrahieren, das von der ersten Eingabeeinrichtung (1) abgegeben worden ist;
  - mit einer zweiten Merkmalsextraktionseinrichtung (20), welche mit der zweiten Eingabeeinheit (2) verbunden ist, um ein zweites Merkmal in jedem der Frequenzbänder aus dem Bezugssignal zu extrahieren, das von der zweiten Eingabeeinrichtung (2) abgegeben worden ist;
  - mit einer Koeffizientenberechnungseinrichtung (40), welche mit der ersten und der zweiten Merkmalsextraktionseinrichtung (10, 20) verbunden ist, um einen Geräuschkoeffizienten in jedem der Frequenzbänder basierend auf dem ersten und dem zweiten Merkmal zu berechnen, wenn keine Sprache des Sprechers vorliegt;
  - mit einer Geräuschannulliereinrichtung (50), welche mit der ersten und zweiten Merkmalsextraktionseinrichtung (10, 20) und der Koeffizientenberechnungseinrichtung (40) verbunden ist, um ein

Sprachmerkmal in jedem der Frequenzbänder basierend auf dem ersten und zweiten Merkmal und dem Geräuschkoeffizienten durch Eliminierung von Geräuschkomponenten aus dem ersten Merkmal abzuleiten;

mit einer ersten Sprachintervall-Feststelleinrichtung (30) zum Feststellen eines generellen Sprachintervalls, welches der Sprache oder Stimme des Sprechers in dem akustischen Signal entspricht, und mit

einer zweiten Sprachintervall-Feststelleinrichtung (60), welche mit der ersten Sprachintervall-Feststelleinrichtung (30) und der Geräuschannuliereinrichtung (50) verbunden ist, um ein Sprachintervall, in jedem der Frequenzbänder in einem Feststellintervall auf Grund des von der Geräuschannuliereinrichtung (50) gelieferten Merkmal festzustellen, wobei das Feststellintervall aus dem generellen, von der ersten Sprachintervall-Feststelleinrichtung festgestellten Intervall und einem vorherbestimmten Intervall gebildet wird, das zu dem generellen Sprachintervall addiert wird.

2. Sprachintervall-Feststelleinheit nach Anspruch 1, dadurch gekennzeichnet, daß die zweite Sprachintervall-Feststelleinrichtung (60) eine Bestimmungseinrichtung aufweist, um zu bestimmen, ob das Sprachmerkmal in jedem der Anzahl Frequenzbänder größer als ein vorherbestimmter Schwellenwert ist oder nicht, wobei, wenn die Bestimmungseinrichtung feststellt, daß das Sprachmerkmal größer als der Schwellenwertpegel ist, die Sprachintervall-Feststelleinrichtung das Sprachintervall in jedem der Anzahl Frequenzbänder feststellt.

3. Sprachintervall-Feststelleinheit nach Anspruch 2, dadurch gekennzeichnet, daß die Sprachintervall-Feststelleinrichtung einen Schwellenwertgenerator hat, um den Schwellenwertpegel auf der Basis des ersten und des zweiten Merkmals zu erzeugen, welche von der ersten und der zweiten Merkmalsextraktionseinrichtung (10, 20) in einer vorherbestimmten früheren Periode unter der Voraussetzung abgegeben worden sind, daß keine Sprache des Sprechers vorliegt.

4. Sprachintervall-Feststelleinheit nach Anspruch 3, dadurch gekennzeichnet, daß der Schwellenwert periodisch aktualisiert wird.

5. Sprachintervall-Feststelleinheit nach Anspruch 1, dadurch gekennzeichnet, daß die Koeffizientenberechnungseinrichtung (40) dafür ausgebildet ist, um den Geräuschkoeffizienten  $k(i)$  entsprechend der folgenden Formel zu berechnen:

$$k(i) = [X(i) + C_1] / [N(i) + C_2]$$

wobei  $X(i)$  das erste Merkmal in jeder der Anzahl Frequenzbänder unter der Bedingung darstellt, daß keine Stimme des Sprechers vorliegt,  $N(i)$  das zweite Merkmal in jeder der Anzahl Frequenzbänder unter der Bedingung darstellt, daß keine Stimme des Sprechers vorliegt, und  $C_1$  und  $C_2$  konstante Zahlen größer als null sind.

6. Sprachintervall-Feststelleinheit nach Anspruch 5, dadurch gekennzeichnet, daß die konstanten Werte  $C_1$  und  $C_2$  einander gleich sind.

7. Sprachintervall-Feststelleinheit nach Anspruch 1, dadurch gekennzeichnet, daß jede der ersten Eingabeeinrichtungen (1) und der zweiten Eingabeeinrichtungen (2) ein Mikrophon aufweist.

8. Sprachintervall-Feststelleinheit nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß die Sprachintervall-Feststelleinrichtung ferner einen Lautsprecher (2s) aufweist, welchem ein Tonsignal von einer externen Einheit zugeführt wird, und daß die zweite Eingabeeinrichtung (2) eine Einrichtung hat, um das von dem Lautsprecher (2s) gelieferte Tonsignal einzugeben.

9. Sprachintervall-Feststelleinheit nach Anspruch 1, dadurch gekennzeichnet, daß die erste Sprachintervall-Feststelleinrichtung (30) dafür ausgebildet ist, um das generelle Sprachintervall festzustellen, falls eine Leistung einer Sprache, welche aus den Sprechmerkmalen in allen der Anzahl Frequenzbänder erhalten wird, welche von der Geräuschannuliereinrichtung zugeführt worden sind, größer als ein vorherbestimmter Schwellenwert ist.

10. Sprachintervall-Feststelleinheit nach Anspruch 1, dadurch gekennzeichnet, daß die erste Sprachintervall-Feststelleinrichtung (30) dafür ausgebildet ist, um das generelle Sprachintervall festzustellen, falls eine Differenz zwischen einer Gesamtleistung des Tonsignals und einer Gesamtleistung des Bezugssignals größer als ein vorherbestimmter Schwellenwert ist, die Gesamtleistung des Tonsignals eine Gesamtsumme der ersten Merkmale in allen der Anzahl Frequenzbänder ist, welche von der ersten Merkmalsextraktionseinrichtung (10) geliefert worden ist, und die Gesamtleistung des Bezugssignals eine Gesamtsumme der zweiten Merkmale in allen der Anzahl Frequenzbänder ist, welche von der zweiten Merkmalsextraktionseinrichtung (20) geliefert worden sind.

11. Sprachintervall-Feststelleinheit nach einem der vorhergehenden Ansprüche, gekennzeichnet durch deren Verwendung in einem Spracherkennungssystem, um Sprache auf der Grundlage einer Stimme eines Sprechers zu erkennen, wobei das Spracherkennungssystem aufweist:

einen Mustergenerator, um ein Eingabemuster, das der Stimme des Sprechers entspricht, auf der Grundlage von entsprechenden Sprachmerkmalen, welche von der Geräuschannuliereinrichtung (50) erhalten werden, in entsprechenden Sprachintervallen zu erzeugen, welche mittels der zweiten Sprachintervall-Feststelleinrichtung (60) festgestellt worden sind,

eine Speichereinrichtung (80) zum Speichern von Referenzmustern, und

eine Erkennungseinrichtung (90), um Sprache auf der Grundlage des eingegebenen Musters, das durch den Mustergenerator erzeugt worden ist, und auf der Grundlage der Referenzmuster zu erkennen, die in der Speichereinrichtung (80) gespeichert sind.

12. Sprachintervall-Feststelleinheit nach Anspruch 4 oder 10, dadurch gekennzeichnet, daß die Geräuschannuliereinrichtung (50) dafür ausgebildet ist, um das Merkmal  $S(i)$  der Sprache entsprechend der folgenden Formel zu berechnen:

$$S(i) = X(i) - k(i) N(i).$$

Hierzu 4 Seite(n) Zeichnungen

FIG. 1

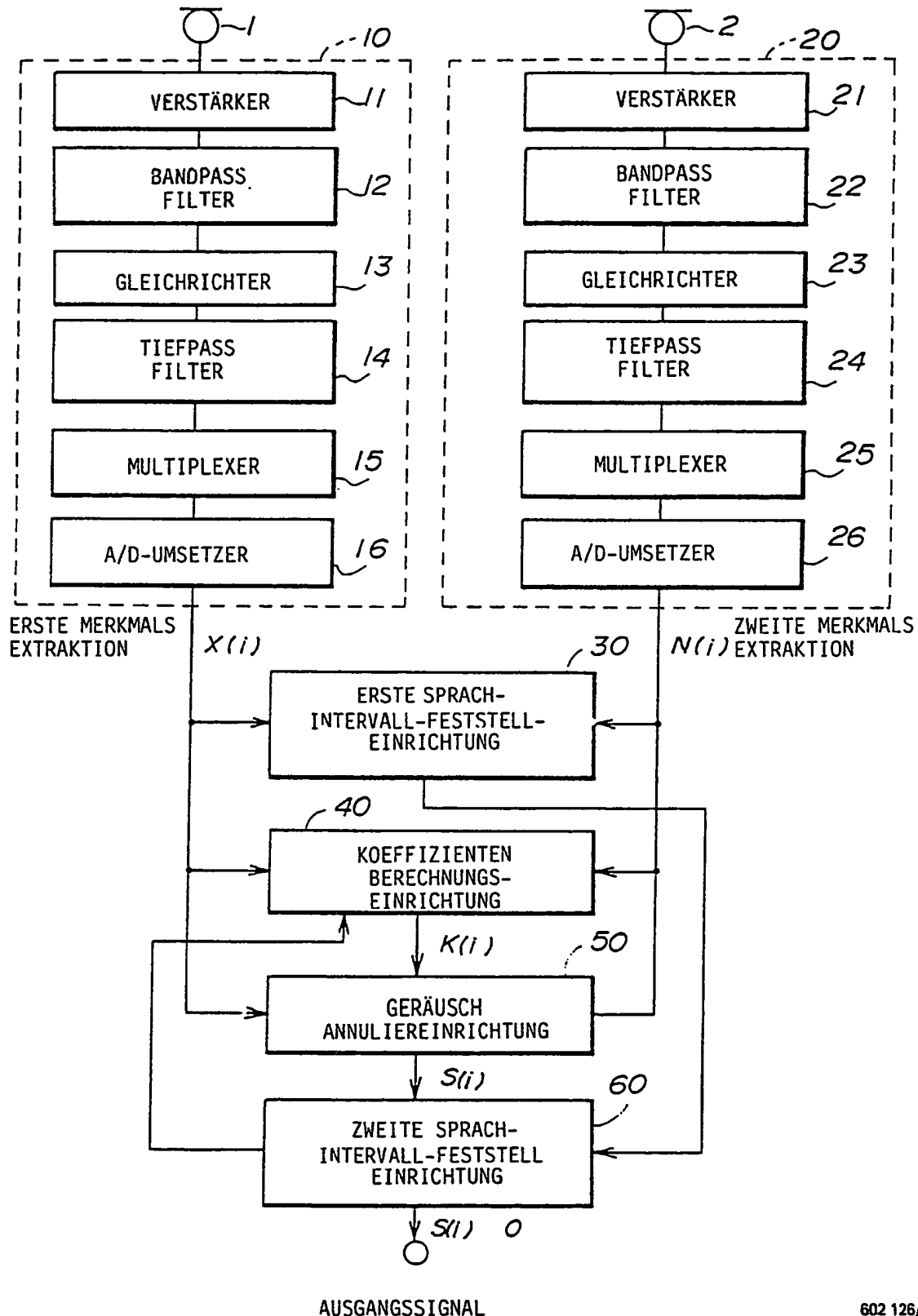




FIG.2

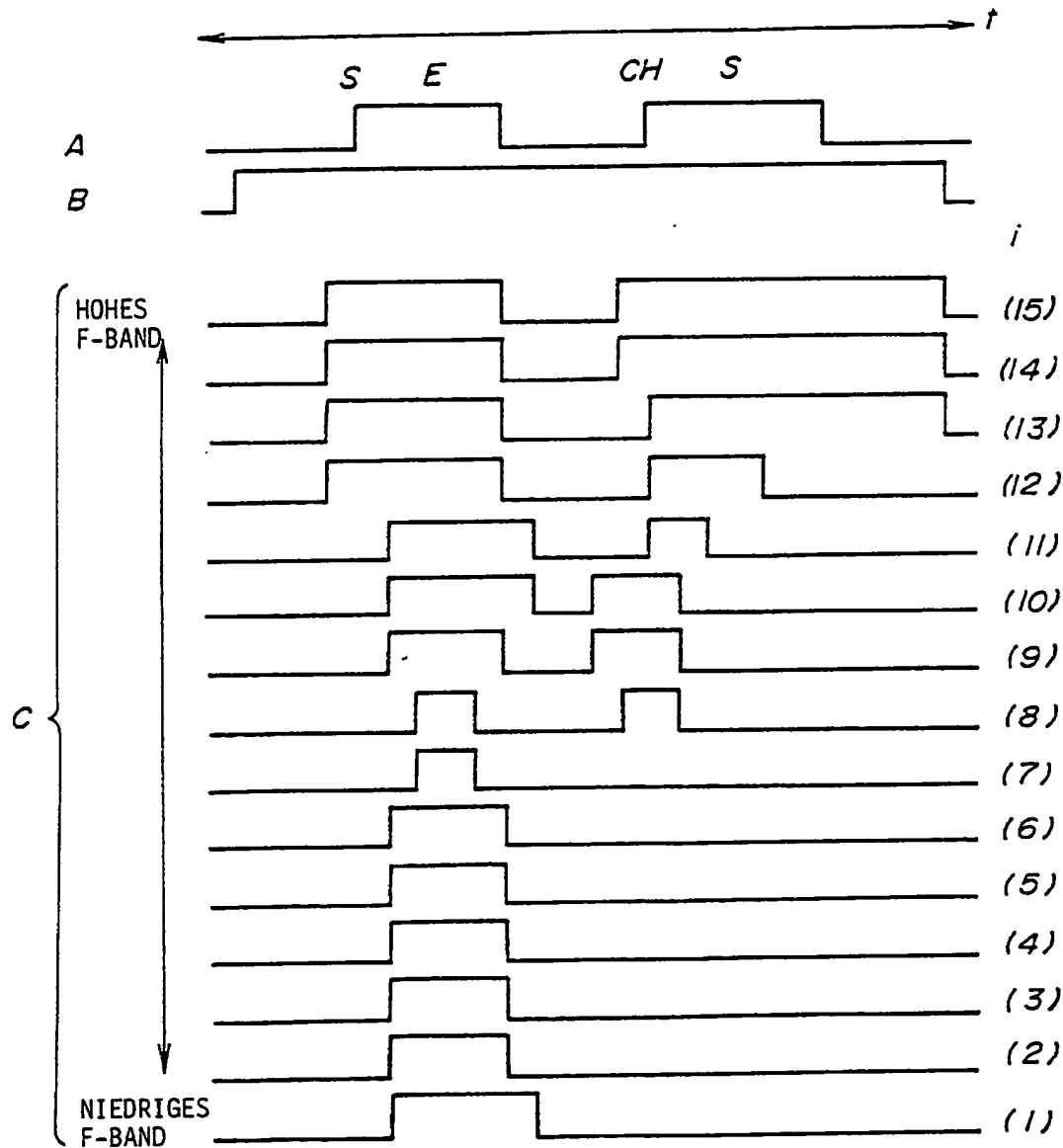


FIG. 3

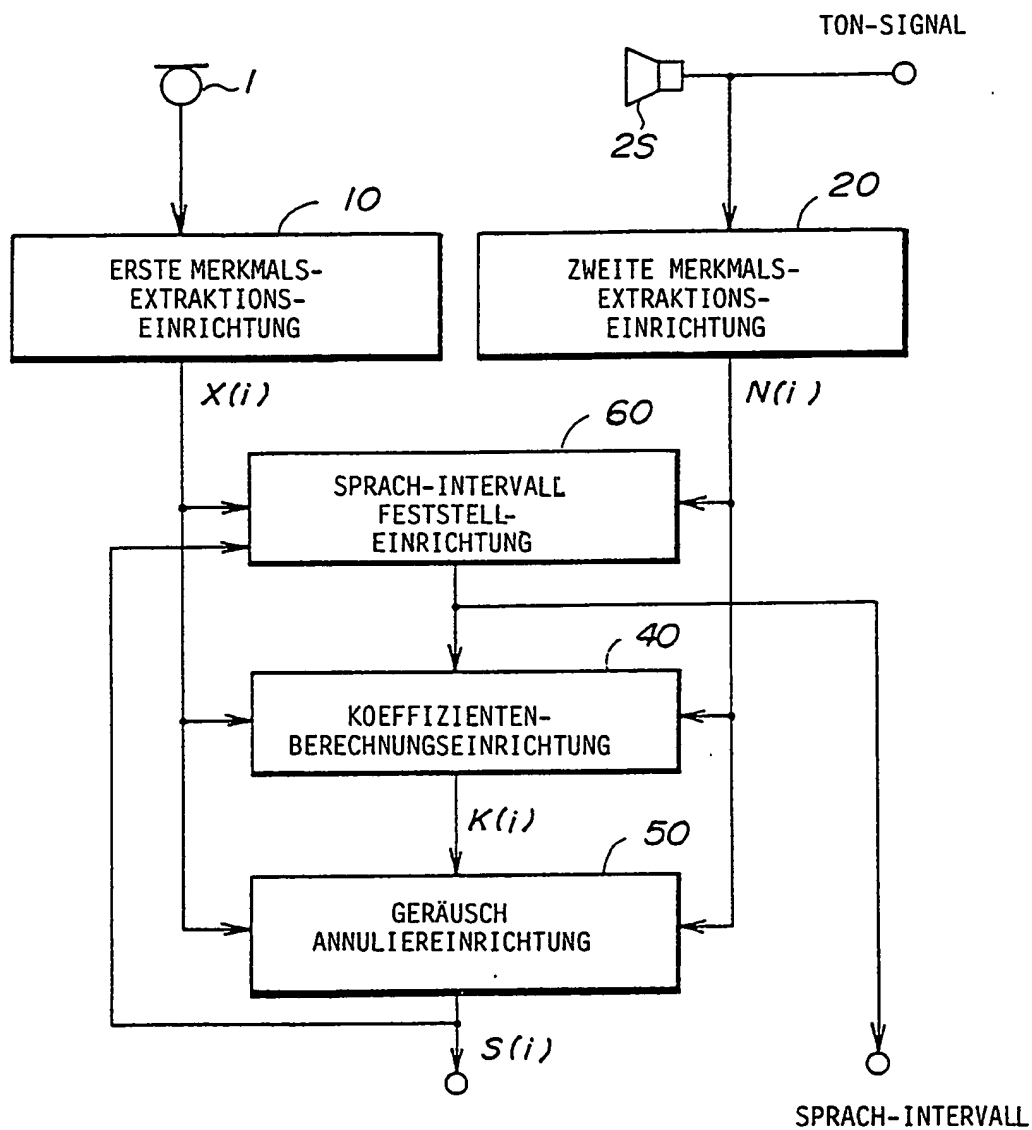


FIG. 4

